

# **Comparison of Maryland Biological Stream Survey (MBSS) and Maryland Stream Waders (MSW) Data and Assessments: Completion of a Statewide Comparison**

*Prepared for:*

**Dan Boward  
Maryland Department of Natural Resources  
Monitoring and Non Tidal Assessment Division  
580 Taylor Avenue, C-2  
Annapolis, Maryland 21401**

*Prepared by:*

**Erik W. Leppo  
And  
Michael J. Paul**

**Tetra Tech, Inc.  
Center for Ecological Sciences  
400 Red Brook Blvd., Suite 200  
Owings Mills, Maryland 21117-5159**

**April 23, 2009**

**(DRAFT FINAL)**

## Introduction

This project continues work begun two years ago to compare data collected and assessments made as part of the volunteer Maryland Stream Waders with those of the Maryland Biological Stream Survey programs (Stribling and Leppo 2007). That project was an effort to objectively define the acceptability of Maryland Stream Waders (MSW) data and assessment results for combining with those produced by the non-volunteer Maryland Biological Stream Survey (MBSS). Although addressing the question of comparability between these two programs was a statewide need, the initial report was focused on fifteen eastern counties associated with Maryland's Coastal Zone Management Program only. This analysis focused on comparisons across the entire state.

Over the last several years there has been a broad effort to expand biological monitoring and assessment, both in terms of increasing an agency's capacity for performing assessments, and in combining monitoring results and datasets from multiple programs to have greater spatial and temporal density of assessments. This has been a concern at federal levels associated with national surveys (US Environmental Protection Agency and US Geological Survey), and by states such as Maryland, whose interest it is to more fully utilize assessments produced by local government (counties and municipalities) and volunteer citizens' organizations. Combining datasets from different programs requires definition of comparability, characterizing data quality in terms of acceptability, and evaluating the effect of hybrid datasets on ultimate uses of the data.

The objective of this study was to determine the level of comparability between bioassessment results collected by professional staff (Maryland Biological Stream Survey [MBSS]) and a volunteers of the Maryland Stream Waders (MSW), and further, to examine it at various scales. We examined the differences of narrative assessment ratings (impaired/non-impaired) of the 2 programs at different spatial scales, as well as the field sampling precision of the programs at each scale.

## Program and data descriptions

The Maryland Biological Stream Survey (MBSS) is a state-wide biological monitoring and assessment program administered by the Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division (MDNR-MANTA). The Maryland Stream Waders (MSW) program is a citizen-based volunteer monitoring program, also sampling a statewide network of stream sites, and also administered by MDNR-MANTA. Sample reaches are randomly-selected (stratified, probability-based) for the MBSS program; for the MSW, volunteers are given individual Maryland 8-digit watersheds by MDNR program managers, and then allowed to choose the specific reaches that they sample. They are, however, given guidelines to ensure that safety is maintained, and that they sample upstream of bridge crossings. Volunteers are also encouraged to avoid always sampling in what they consider to be the highest quality streams or the lowest quality streams.

The MBSS samples benthic macroinvertebrates, fish, physical habitat quality, and selected water chemistry used for evaluating the ecological condition of state water resources, particularly first through third order, Wadeable streams. For benthic macroinvertebrate samples, the gear and methods they use include a long-handled D-frame net with 595µ mesh openings, with which multiple productive habitats are sampled from a 75 meter reach. At a randomly selected subset of sites, duplicate samples are taken by the same field crew, in the same reach, and in the same manner. Sample preservation is in approximately 80% ethanol for transport to the Maryland DNR laboratory, where samples are sorted and subsampled to approximately 100 organisms. The (sub)samples are identified primarily to genus level, including midges (Diptera: Chironomidae) and worms (Annelida: Oligochaeta).

The MSW program samples for benthic macroinvertebrates in exactly the same way as the MBSS, just using volunteers instead of MDNR staff. MSW volunteers go through field-training and testing by MBSS-experienced personnel prior to sampling. Samples are preserved in the same manner, delivered to the MDNR laboratory for sorting and subsampling; taxonomy is performed by MDNR laboratory taxonomic staff. Duplicate samples (repeated sample pairs) are taken for MSW samples by MDNR staff, at the same locations, and no more than 2 weeks later.

Since 1994, there have been over 2,000 samples collected by the MBSS statewide. In their sampling, beginning in 1990, the MSW program has produced nearly twice as many samples statewide. Similarly, the MSW has collected repeated samples nearly twice as frequently as the MBSS.

Taxonomic QC for the MDNR laboratory is based on whole-sample reidentification of a randomly selected subset of samples, and documentation of taxonomic precision. Recent analyses have shown that taxonomic precision for the MBSS dataset is approximately 15% at the genus level (*unpublished data*). Although QC analyses have not been done at family level, logically, the error rate would be expected to be substantially lower. Genus level data are those for which most of these comparability analyses have been done.

Samples included from the two programs in this study were collected in non-tidal Maryland streams from 2000-2007. MBSS samples are included if designated as probability sites, targeted sites were excluded. There are a total of 5,906 samples included in this analysis from both MBSS and MSW (Table 1).

The two programs used similar sampling gear, the same sampling methods, laboratory sample sorting, and taxonomic lab procedures. Besides the differences in field collection personnel the main difference between these programs was in the level of taxonomic identifications and the selection of the sampling locations (Table 2). MBSS followed a rotating basing plan for its sampling design. In most cases MSW sampled the same watersheds in the same years as MBSS. Whereas MBSS samples were selected as part of a stratified random sampling plan, MSW sites were selected by the volunteers from the MBSS provided watersheds based mainly on access. Samples from both programs were selected and identified by the same MBSS staff

in the MBSS lab facility. MBSS samples were identified to genus and MSW samples were only identified to family.

Samples were scored using the appropriate MBSS benthic index of biological integrity (BIBI). MBSS data were scored using a genus level BIBI (Southerland et al. 2005) and MSW data using the family level BIBI (Stribling et al. 1998). Both BIBI's were scored on a 1 – 5 scale: 1.0-1.9 = very poor, 2.0-2.9 = poor, 3.0-3.9 = fair, and 4.0-5.0 = good. Scores < 3 were considered degraded.

## Methods

The methods used were comparable to what were used in the previous coastal zone comparison. We looked at three different aspects of comparability: shared taxonomic composition between repeated samples, field sampling precision (using final index scores), and assessment precision (using final assessment categories).

Within-method comparisons were made between primary and duplicate samples at the site level before comparisons were made at larger scales. The MBSS and MSW data could not be compared directly due to differences in level of taxonomic effort and differences in the BIBI that was used. However, it was possible to compare the final assessment derived using samples from each program. It was also possible to compare the expected results of the MSW index by collapsing the MBSS genus level data to family level and then using the family level BIBI. MBSS samples analyzed with the MBSS genus BIBI are referred to as “MBSS (genus)”. MBSS samples analyzed with the MSW family BIBI are referred to as “MBSS (family)”. MSW samples are referred to as “MSW (family)”. This analysis showed that the same samples analyzed by either the family or genus level BIBI were nearly identical. The MSW data, while having a mean similar to that of the MBSS data adjusted to family level, was more variable (Table 3).

Comparisons between the final assessments of the two programs were made on multiple study unit levels: State-wide, MD HUC (MDE) 6 digit, MD HUC (MDE) 8 digit, MD HUC (DNR) 12 digit, Stream-reach (NHD+). In addition, site specific comparisons were made with duplicate samples within each method. For comparisons only study units with more than 2 samples from each program were used. The results of each program's primary samples are discussed below. Duplicates were only considered in the site-specific section.

Taxonomic composition of repeated sample pairs—We evaluate composition between two samples as percent of taxa shared between the two samples, and the percent of taxa each sample contains that is unique to it (relative to the other sample). Combining these percentages across all sample pairs (MBSS x MBSS, genus and family level) and (MSW x MBSS, family level), and present median values. A lower percentage of shared taxa between two samples is expected the more dissimilar or inconsistent field sampling is.

Field sampling precision—We calculated several precision terms for field sampling using final index scores from the sample pairs. Biological index scores are used for calculations. The precision terms used are root mean square error (RMSE), coefficient of variability (CV), and confidence interval (CI) (equivalent to detectable difference [DD]). Formulas used in these

calculations can be found in Flotemersch et al. (2006), and Stribling et al. (2008). For all of these terms, better precision (repeatability, consistency) of the dataset overall is reflected by low values.

Assessment precision—The precision of final assessment is reflected as the rate of agreement (or disagreement) between each sample of a duplicate pair and how they translate to either non-degraded (good or fair) or degraded (poor and very poor). Scoring categories are based on the MBSS IBI (Stribling et al. 1998, Southerland et al. 2005). Additionally, we have looked at the sample pairs in terms of the proportion of times they differ by 0, 1, 2, or 3 assessment categories (good, fair, poor, or very poor). All were done for the MBSS sample pairs as well as the MSW-MBSS sample pairs.

Issues - Several issues arose during data preparation and analysis. These are issues that may be important to MBSS and their dataset and in merging MBSS-MSW in the future. These are noted in Appendix 1 and MBSS is encouraged to consider these in moving forward.

## **Results and Discussion**

### **Comparisons**

#### ***State-wide***

All primary samples collected by MBSS and MSW were aggregated to calculate an assessment for the entire state (Table 4). Both programs give a narrative assessment of poor.

#### ***HUC06***

Of the 20 watersheds in Maryland at the 6 digit level (as defined by MDE) 19 have been sampled. The 20<sup>th</sup> is tidal and is not part of the MBSS study (this watershed includes the Chesapeake Bay proper). MBSS has sampled 18 of the 19 sampleable watersheds and MSW 19. There are 18 watersheds with at least 2 samples from each program. Overall, there was general agreement between the narrative assessments of the 2 programs (Table 5) but the majority (78%) was off by 1 category. However, no HUC06 watersheds were assessed as >1 category different. There was 72% agreement in degraded/non-degraded assessment between the two methods (Figure 1).

#### ***HUC08***

There were 138 watersheds in Maryland defined by MDE at the 8-digit level. MBSS has sampled 122 and MSW has also sampled 122. There were 106 with at least 2 samples from each program. There was general agreement between the narrative assessments of the 2 programs (Table 6) but 57% were different by 1 category and 5% were different by >1 category. There was 66% agreement in degraded / non-degraded assessment between the two methods (Figure 2).

#### ***HUC12***

There were a total of 1,124 watersheds in Maryland defined by DNR at the 12-digit level. MBSS sampled 565 and MSW sampled 788. There were 222 with at least 2 samples from

each program. There was general agreement between the MBSS and MSW narrative ratings (Table 7) though almost 50% were rated 1 category different. At the watershed level, HUC12 watershed assessments exhibited the most agreement (44% agreement vs 39% for HUC08 and 22% for HUC06). There was 66% agreement in degraded / non-degraded assessment between the two methods (Figure 3).

### **Stream-reach**

There were a total of 13,345 stream reaches in Maryland as defined in the National Hydrography Dataset plus (NHD+) (<http://www.horizon-systems.com/nhdplus/documentation.php>). A stream reach is defined as an unbroken section stream between nodes in the NHD+ dataset. This included all rivers and streams (wadeable/non-wadeable and tidal/non-tidal). MBSS has sampled 1,385 reaches and MSW 2,446 reaches. A combined total of 3,198 unique stream reaches were sampled by the two programs but only 39 reaches had a minimum of 2 samples from each program required to do the comparison. It was noted that the most number of samples from both programs on the same reach was 5 samples. There was general agreement in the narrative assessments between the programs (Table 8) at the stream reach study unit. There were no stream reaches with a narrative difference of 3 categories but 18% (7) were assessed as 2 categories different. There was, again, 66% agreement in degraded / non-degraded assessment between the two methods (Figure 4).

### **Sample Taxonomic Composition**

MBSS and MSW duplicates were examined for potential bias in the taxa that were collected by each program, and for its effect on assessment scores and narratives. Shared taxa are taxa that were common between the primary and duplicate sample pairs. Unique taxa are those taxa found in only one of the sample pairs. Data were summarized as means and standard deviations across all pairs, state-wide. To compare on a common scale and to alleviate problems of comparing taxa rich and taxa poor samples, percent of taxa in sample was used rather than number of taxa. For the MBSS samples a similar number of taxa were unique to both samples in each pair whether genus (Figure 5) or family (Figure 6) level identifications were used. This was also the case with the MSW samples (Figure 7) even though the samples were collected by different teams.

### **Sample Collection Effort**

In addition to what was collected (sample composition), sample collection effort was examined to evaluate bias. Sample collection effort was defined as the amount of material collected and the portion needed to obtain the minimum 100 organisms for a sample. The lab sorting personnel and procedures were the same between the programs (i.e., the same MBSS lab processes samples using the same methods for both programs). MBSS samples (primary and duplicate) required, on average, 13 grids to reach the target 100 organisms. In contrast, MSW primary samples required, on average, 34 grids to get to the minimum number of organisms. The MSW duplicate samples (collected by MBSS staff) needed, on average, 20 grids. This was higher than MBSS program samples but lower than MSW volunteer collected samples (Table 9). This difference could be attributed to the quality of the samples and the available habitat at

each site as well as the overall amount of material collected by the personnel of the two programs. Additional study would be necessary to identify the reasons for these differences.

### **Assessment Precision**

Assessment precision is the agreement in assessment (degraded/non-degraded status) between the primary and duplicate samples within a program. Agreement was high among MBSS (genus) samples (86%) (Table 10) and increased to 92% using the family level BIBI. The MSW samples exhibited 81% agreement.

Narrative assessment differences between primary and duplicate samples were comparable. There were no samples with differences of 3 assessment categories and the majority of samples for each program received the same narrative assessment (Table 11).

### **Summary and Recommendations**

In general, both MSW and MBSS samples reflected similar taxonomic diversity. This is encouraging in that the basis for these assessments is ostensibly taxonomic diversity and the relative abundance of invertebrates. These appear to be comparable between the two programs statewide and lend support to the notion that they reflect similar judgments regarding biological condition. In addition, the precision of field sampling was similar among replicates, indicating a similar level of field method error. This means that the relative error around assessments due to field sample error should be comparable. Moreover, the precision of assessments in the statewide comparison did not change from the earlier estimates using coastal zone data alone (~83-85%). Lastly, and most importantly, there was general agreement in assessment results, although a consistent bias appeared where by MSW scores were consistently higher than MBSS scores (regression lines in Figures 1-4 are above the 1:1 line) and the agreement appeared to increase at smaller scales. Where narrative assessments differed, it was generally by only 1 narrative category, except at the smallest scales where a few large differences occurred.

### **Items for Further Study**

Several items emerged through analysis and consultation as important for further study. These are items that we recommend MBSS may want to pursue in the future.

*Compare percent of sample analyzed among MSW primary duplicate samples.* The percent of sample analyzed varied quite a bit across MSW samples and in comparison to MBSS samples. Comparing duplicate MSW samples would allow an analysis of the extent of systematic variability in sample percent.

*Test sensitivity to stressors (data unavailable).* No stressor data were available for MSW sites. It may be worthwhile to test whether and to which stressors MSW data respond and how it compares to stressor response of MBSS indices. Both could be plotted against similar stressors together to compare the slope of these responses as well.

*Evaluate effect on HUC 8 assessments of non-random site selection of MSW sites.* It was outside the purview of this project to compare how the difference in study design affected



overall MBSS HUC 8 assessments. Our comparisons were made assuming comparability. But, this issue likely deserved more attention.

*Break data into smaller subsets (by year or bioregion) in order to ascertain where difference may be occurring.* Differences were observed among the two programs. This could be due to many sources, including sampler bias among others. Further refinement of the data would be needed to tease these sources apart.

*Further examine differences in sample collection effort between MBSS and MSW personnel.* Similarly, there was reason to expect that differences among the two programs may be due to differences in effort. This analysis would require more specific effort focused on sampler level effort.

### **Literature Cited**

Flotemersch, J.E., J.B. Stribling, and M.J. Paul. 2006. Concepts and Approaches for the Bioassessment of Non-wadeable Streams and Rivers. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C. EPA/600/R-06/127.

Southerland, M. T., G. M. Rogers, M. J. Kline, R. P. Morgan, D. M. Boward, P. F. Kazyak, R. J. Klauda, and S. A. Stranko. 2005. New Biological Indicators to Better Assess the Condition of Maryland Streams. Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment, Annapolis, Maryland.

Stribling, J.B., K.L. Pavlik, S.M. Holdsworth, and E.W. Leppo. 2008. Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *Journal of the North American Benthological Society*. 27(4):906-919.

Stribling, J. and E.W. Leppo. 2007. Comparison of Maryland Biological Stream Survey (MBSS) and Maryland Stream Waders (MSW) Data and Assessments. Prepared by Tetra Tech, Inc. for the Maryland Coastal Zone Management Program.

Stribling, J.B., B.K. Jessup, J.S. White, D. Boward, and M. Hurd. 1998. Development of a Benthic Index of Biotic Integrity for Maryland Streams. Chesapeake Bay and Watershed Programs. Monitoring and Non-Tidal Assessment, Annapolis, Maryland. CBWP-MANTA-EA-98-3.



**Table 1.** Number of samples included in study, by program.

<b>Sample Type</b>	<b>MBSS</b>	<b>MSW</b>	<b>Total</b>
Primary	1192	4497	5689
Duplicate	66	151	217
TOTAL	1258	4648	5906

**Table 2.** Maryland Biological Stream Survey (MBSS) and Maryland Stream Waders (MSW) program highlights.

<b>Feature</b>	<b>MBSS</b>	<b>MSW</b>
Site selection	Stratified random selection	Selected by sampler (non random) within MBSS provided watersheds
Sample Collection	Spring; D-net, multi-habitat, 20 ft <sup>2</sup> , field preserved	Spring; D-net, multi-habitat, 20 ft <sup>2</sup> , field preserved
Field Crews	Professional	Volunteer
Field Crew Training	Yes	Yes
Sample Collection Time Period	Spring	Spring
Sample Processing	MBSS Lab – 100 organism	MBSS Lab – 100 organism
Taxonomy	MBSS Lab – Genus	MBSS Lab - Family
Duplicate Samples	Collected the same day by the same crew at the same reach	Collected a minimum of 2 weeks later by a different crew (MBSS personnel instead of volunteers) at the same reach

**Table 3.** Field sampling precision of each method. N = number of records. MSE = mean square error. RMSE = root mean square error. CV = coefficient of variability. CI90 = 90% confidence intervals. Definition of terms can be found in Flotemersch et al. 2006.

<b>Program (ID Level)</b>	<b>N</b>	<b>MSE</b>	<b>RMSE</b>	<b>Mean</b>	<b>CV</b>	<b>CI90</b>
MBSS (genus)	66	0.09	0.29	2.71	10.78	0.48
MBSS (family)	66	0.09	0.29	2.24	13.11	0.48
MSW (family)	145	0.25	0.50	2.27	22.14	0.83

**Table 4.** MBSS and MSW BIBI statewide assessments.

<b>Program (ID Level)</b>	<b>Number</b>	<b>Mean</b>	<b>Median</b>	<b>Standard Deviation</b>	<b>Narrative Assessment</b>	<b>Status</b>
MBSS (genus)	1150	2.77	2.75	1.03	Poor	Degraded
MBSS (family)	1150	2.34	2.43	0.80	Poor	Degraded
MSW (family)	4234	2.24	2.14	0.87	Poor	Degraded

**Table 5.** Narrative assessment difference at the MDE06 watershed level. Narrative difference is the number of categories different in the final BIBI assessments.

<b>Program (ID Level)</b>	<b>N</b>	<b>Narrative Difference</b>			
		<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
MBSS (genus) – MSW (family)	18	22%	78%	0%	0%

**Table 6.** Narrative assessment difference at the MDE08 watershed level. Narrative difference is the number of categories different in the final BIBI assessments.

<b>Program (ID Level)</b>	<b>N</b>	<b>Narrative Difference</b>			
		<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
MBSS (genus) – MSW (family)	106	39%	57%	4%	1%

**Table 7.** Narrative assessment difference at the DNR12 watershed level. Narrative difference is the number of categories different in the final BIBI assessments.

Program (ID Level)	N	Narrative Difference			
		0	1	2	3
MBSS (genus) – MSW (family)	222	44%	48%	7%	1%

**Table 8.** Narrative assessment difference at the reach level. Narrative difference is the number of categories different in the final BIBI assessments.

Program (ID Level)	N	Narrative Difference			
		0	1	2	3
MBSS (genus) – MSW (family)	39	44%	38%	18%	0%

**Table 9.** Program sample collection laboratory subsampling effort (i.e., the number of grids sorted to obtain 100 organism subsample).

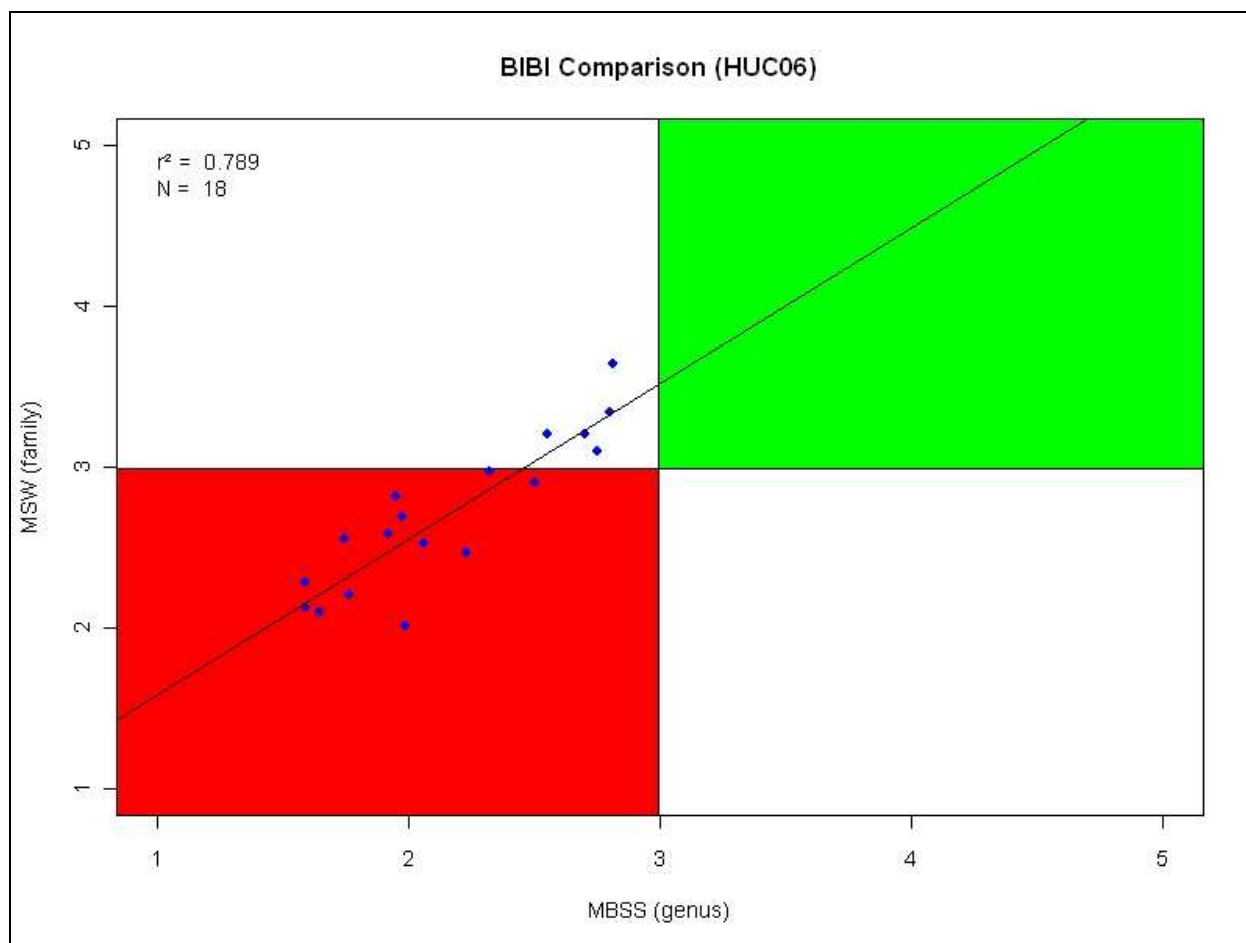
Program	Sample Type	Number	Mean	Standard Deviation
MBSS	Primary	66	12.9	8.4
	Duplicate		13.5	13.0
	<i>difference</i>		5.5	8.0
MSW	Primary	116	34.0	35.0
	Duplicate		19.9	28.8
	<i>difference</i>		22.2	27.4

**Table 10.** Narrative assessment agreement within programs comparing primary and duplicate samples.

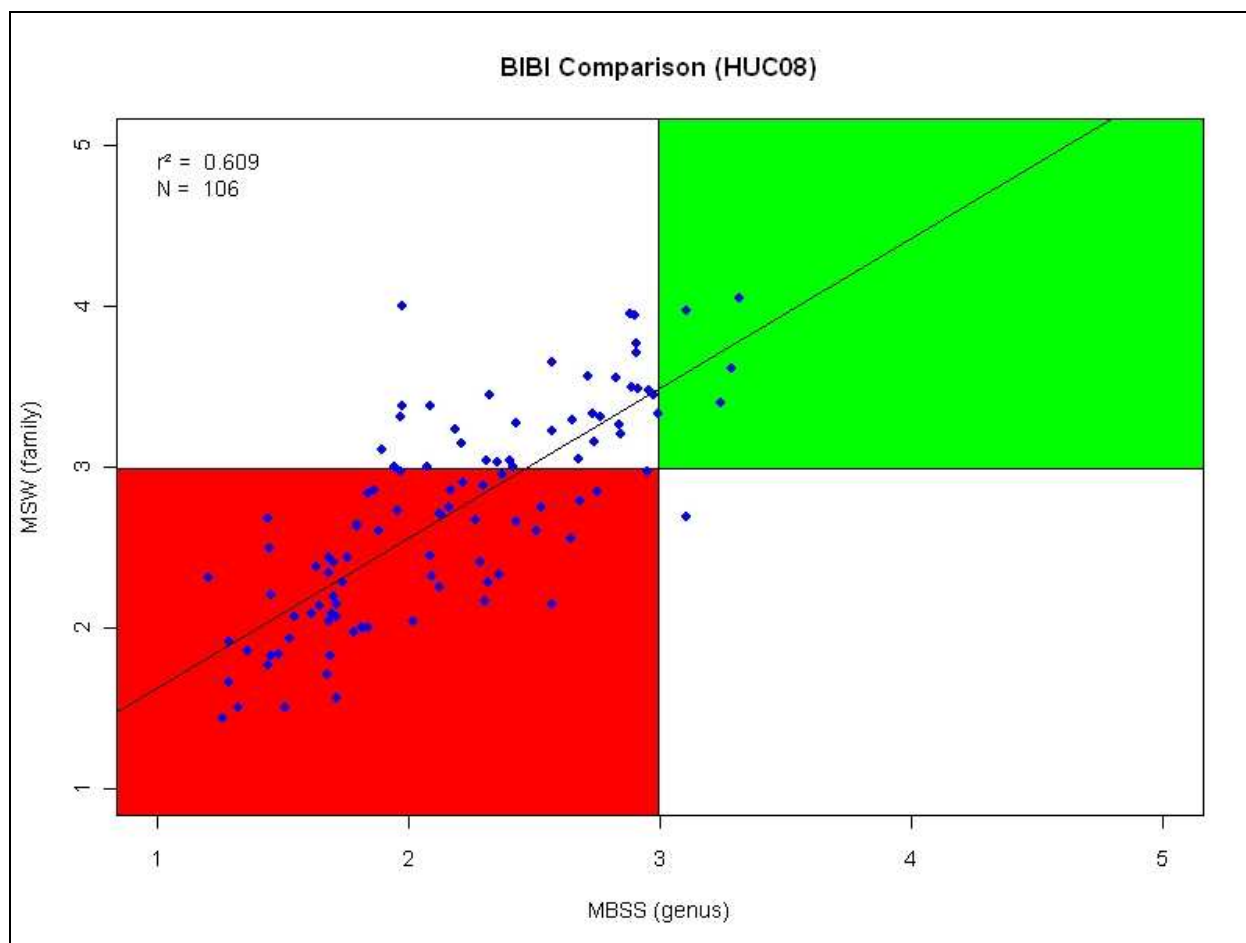
Program (ID Level)	Total N	Primary Sample Assessment	Duplicate Sample Assessment		Percent Agreement	
			Non-Degraded	Degraded	Category	Overall
MBSS (genus)	66	Non-Degraded	<b>24</b>	4	85.7	86.4
		Degraded	5	<b>33</b>	86.8	
MBSS (family)	66	Non-Degraded	<b>13</b>	1	92.9	92.4
		Degraded	4	<b>48</b>	92.3	
MSW (family)	145	Non-Degraded	<b>30</b>	10	75.0	81.4
		Degraded	17	<b>88</b>	83.8	

**Table 11.** Narrative assessment difference at the site level (between primary and duplicate samples).

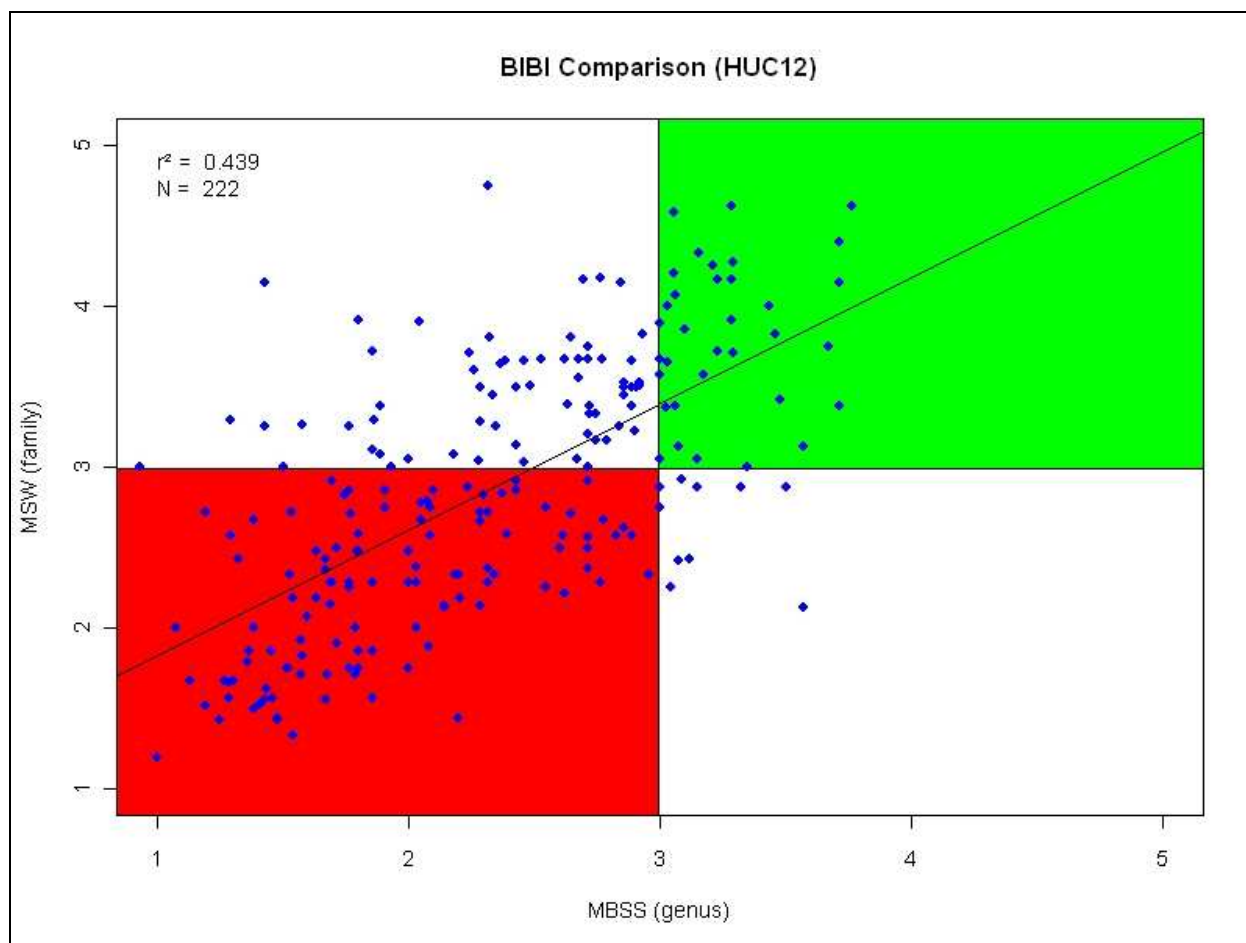
Program (ID Level)	N	Narrative Difference			
		0	1	2	3
MBSS (genus)	66	62%	35%	3%	0%
MBSS (family)	66	74%	24%	2%	0%
MSW (family)	145	67%	29%	4%	0%



**Figure 1.** Comparison of MBSS and MSW BIBI scores at the MDE 6 digit HUC level. The shaded areas are where there was agreement between the two programs; green for non-degraded and red for degraded. Each point represents samples within the same sampling unit.

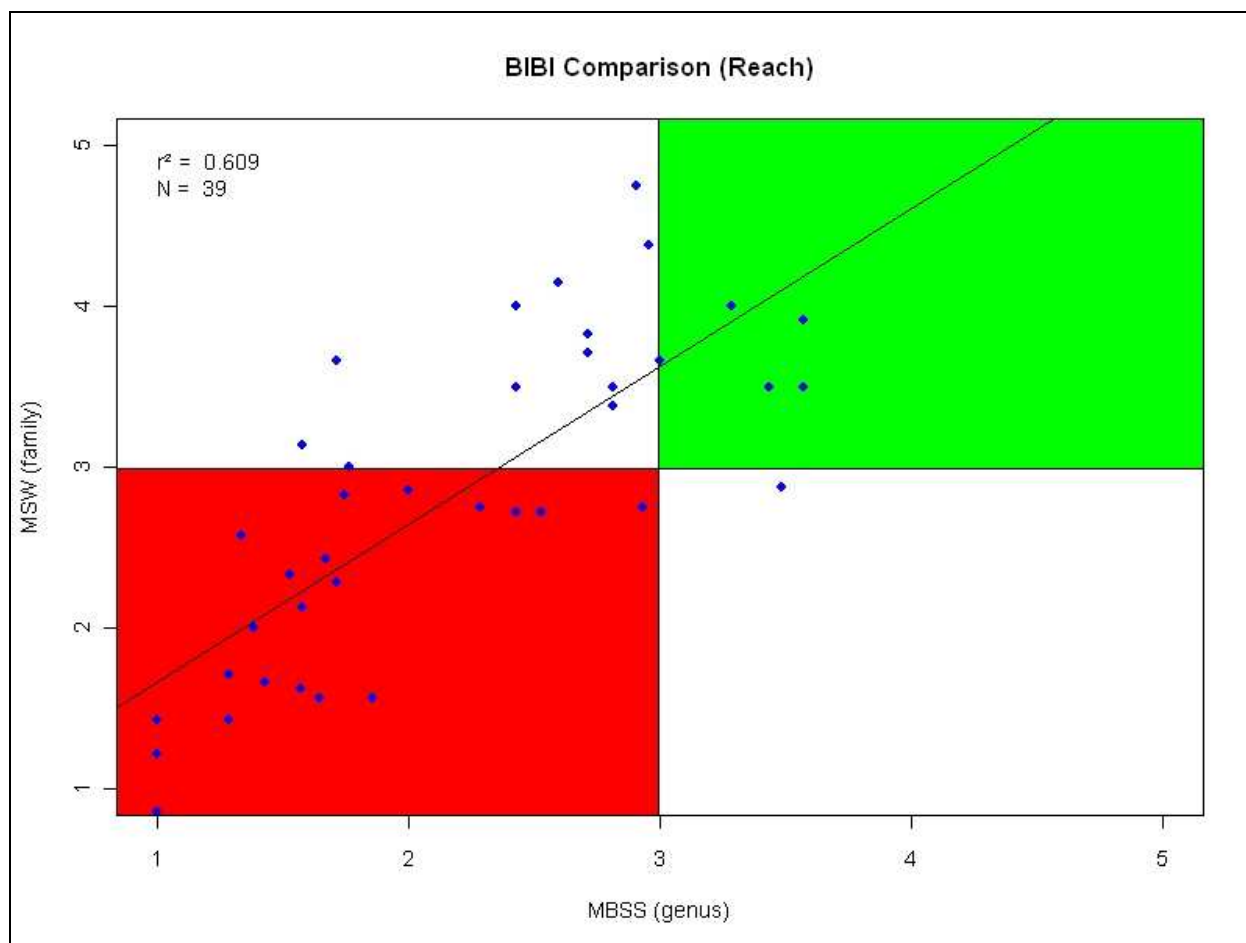


**Figure 2.** Comparison of MBSS and MSW BIBI scores at the MDE 8 digit HUC level. The shaded areas are where there was agreement between the two programs; green for non-degraded and red for degraded. Each point represents samples within the same sampling unit.

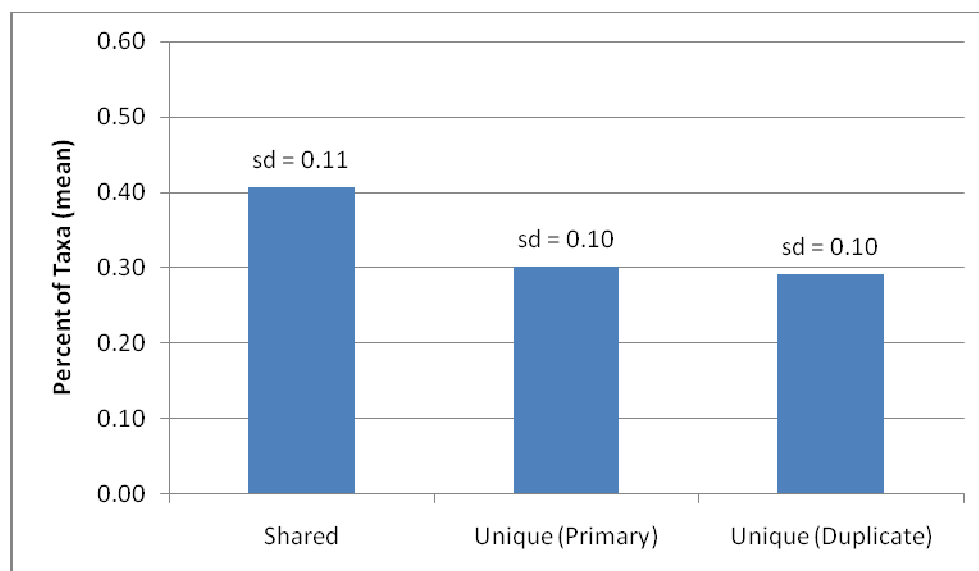


**Figure 3.** Comparison of MBSS and MSW BIBI scores at the DNR 12 digit HUC level. The shaded areas are where there was agreement between the two programs; green for non-degraded and red for degraded. Each point represents samples within the same sampling unit.

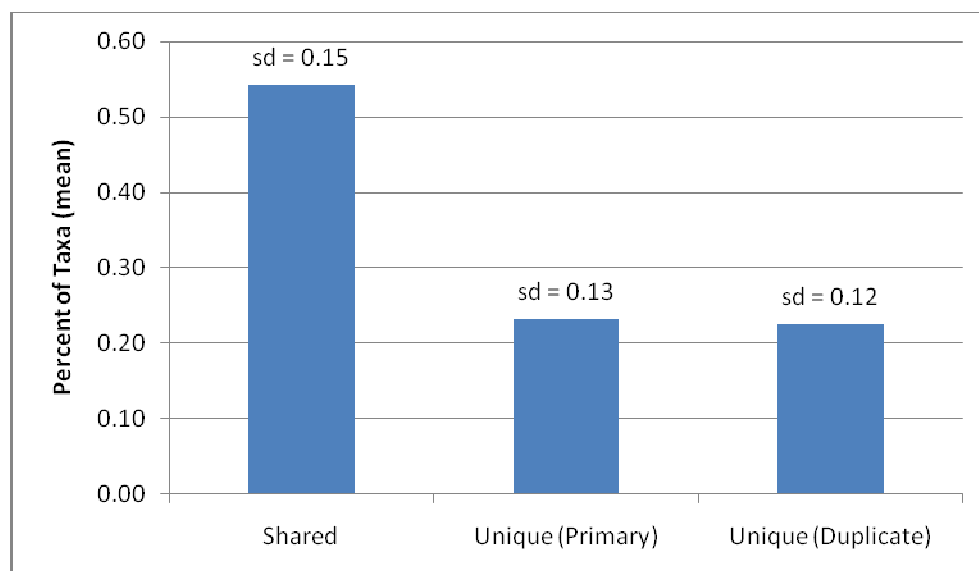




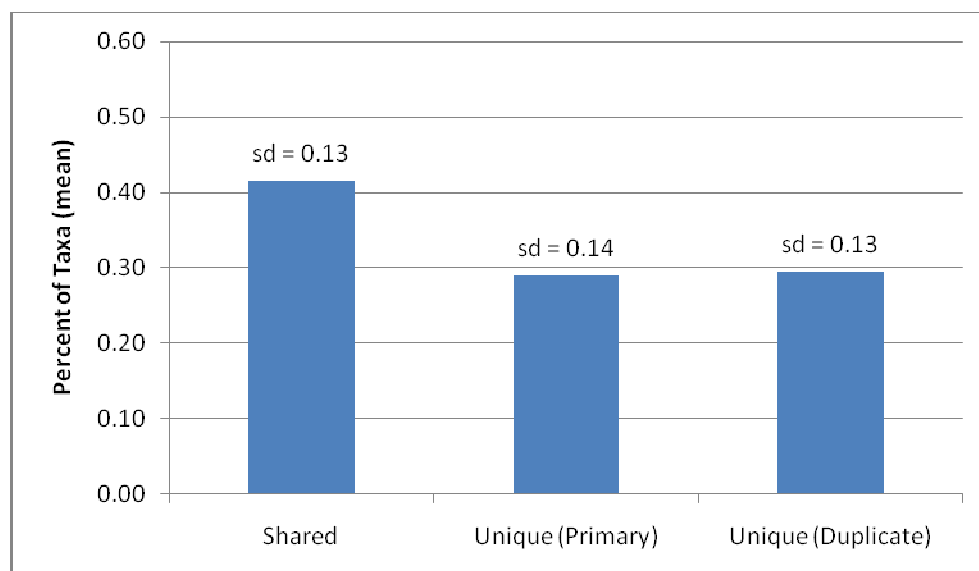
**Figure 4.** Comparison of MBSS and MSW BIBI scores at the reach level. The shaded areas are where there was agreement between the two programs; green for non-degraded and red for degraded. Each point represents samples within the same sampling unit.



**Figure 5.** Taxonomic composition comparison between duplicate sample pairs, MBSS genus, N=66. Each bar represents the mean value between pairs for shared taxa (in both samples) and unique taxa (in only one sample of each pair). sd = standard deviation.



**Figure 6.** Taxonomic composition comparison between duplicate sample pairs, MBSS family, N=66. Each bar represents the mean value between pairs for shared taxa (in both samples) and unique taxa (in only one sample of each pair). sd = standard deviation.



**Figure 7.** Taxonomic composition comparison between duplicate sample pairs, MSW, N=145. Primary collected by volunteers and duplicate collect by professionals (MBSS personnel). Each bar represents the mean value between pairs for shared taxa (in both samples) and unique taxa (in only one sample of each pair). sd = standard deviation.

## **Appendix 1 – Issues involved in data preparation and analysis that required correction**

During the compilation of the data and during the analysis some complications arose with the data. We briefly discuss the major issues and make some recommendations for fixing them with the goal of long term data integrity.

**No primary keys in the data tables were provided that allowed for the creation of database relationships among the tables.** Primary keys should be established that identify the stations as unique points in space. The samples should then be linked on this new primary key. The current identifier is a hybrid of station and sample identifier but does not effectively function as either. For this analysis, we created a station ID from the initial portion of the station/sample identifier. Discrepancies in the format were corrected when obvious.

**Duplicate samples had different identifiers.** This is an extension of the first issue. We changed station identifiers so that primary and duplicate samples could be matched. We changed letter codes to numeric digits for MBSS (which were a mix of 1 character and all 3 characters). We also removed the MSW leading “9” so we could match sites on station ID.

**The MSW site identifier for watershed code was not always 4 digits.** Watershed code was the last 4 digits of DNR12 digit code, so we changed all station identifiers to 4 digit codes with leading zeros to make sites more comparable.

**There were missing data for watershed codes (MDE6, MDE8, DNR12, and ShedCode), so we used ArcMap to apply.** Many sites were missing this information, even more were missing leading “0”s. We used ArcMap to apply the correct watershed codes to these sites.

**The MSW data contained a mix of data types in latitude and longitude fields.** To correct this, we used Northing and Easting.

**The metric sums and IBI scores did not agree in all cases (MSW and MBSS).** We used metric values to recalculate metric scores and then recalculated the BIBI scores and narrative rankings. To do this, we used the master taxa list as provided.

**There was no stream reach information for many sites.** To correct this, we used NHD+ dataset after consultation with client. We used Northing/Easting to match stream reaches with NHD+ locations. Many points were not on blue lines, so we used the nearest match. These stream reach identifications need a more critical review.

**The master taxa list was missing taxa (18) that were included in samples.** These taxa were added to the master taxa list prior to metric calculation along with associated phylogenetic information. ITIS was used as source for this information ([www.its.gov](http://www.its.gov)).